ABSTRACT OF THE DISCLOSURE

"A Method for Discovering Undeclared and Fuzzy Rules in Databases"

A scheme is used to automatically discover algebraic constraints between pairs of columns in relational data. The constraints may be "fuzzy" in that they hold for most, but

5   not all, of the records, and the columns may be in the same table or different tables. The scheme first identifies candidate sets of column value pairs that are likely to satisfy an algebraic constraint. For each candidate, the scheme constructs algebraic constraints by applying statistical histogramming, segmentation, or clustering techniques to samples of column values. In query-optimization mode, the scheme automatically partitions the data

10   into normal and exception records. During subsequent query processing, queries can be modified to incorporate the constraints; the optimizer uses the constraints to identify new, more efficient access paths. The results are then combined with the results of executing the original query against the (small) set of exception records.